Biomedical Microdevices (2019) 21: 37
https://doi.org/10.1007/s10544-019-0392-y

Project honeybee: Clinical applications for wearable biosensors

Leland Hartwell[1] · Heather M. Ross[2,3] · Jeffrey T. La Belle[4,5]

Published online: 1 April 2019

Abstract
Please provide an abstract of 150 to 250 words. The abstract should not contain any undefined abbreviations or unspecified references. The Project Honeybee Observational Clinical Trials were 12-month studies designed to validate the use of commercially available ambulatory medical devices costing $50–$300 for clinical applications. Each trial had a patient population of about 15–30 subjects with a broad range of disease types including heart failure, diabetes, sepsis, and Parkinson's disease. Over 30 supported proposals were funded in the 4-year period, as well as the creation of a database of all commercially available devices. Each year a call for proposals was published within ASU and Mayo Clinic Arizona. Proposals were selected for funding by a committee of ASU faculty from engineering, nursing, and exercise physiology departments. The progress of each research trial was monitored through monthly colloquia with the nursing, biomedical engineering, computer science, and nutrition graduate research assistants, to discuss the challenges and opportunities arising with each research trial. PIs were required to report on study progress 6 months into the trial period and 3 months following the conclusion of the 12-month project. The project was very successful in meeting our goals of testing consumer wearable devices on patients for a variety of conditions across a variety of clinical settings in the greater Phoenix community. The following clinical sites participated in one or more of these clinical trials: Adelante Healthcare, Arizona Arrhythmia Consultants, Arizona Cardiology Group, Banner University Medical Center, Barrow Neurological Institute, Honor Health, Mayo Clinic, and St Joseph's Hospital. A total of 12 ASU faculty and 39 clinicians participated.

Keywords Project honeybee · Wearables · Sensors · Clinical medicine

1 Introduction

Modern airline travel is widely recognized as one of the safest modes of transportation worldwide. This is due, in part, to the advent of continuous sensory monitoring (i.e. sound, vibration, temperature, and pressure) of mechanical aircraft components (Reliability Centered Maintenance) in order to detect parts needing replacement prior to mechanical failure. Similarly, improving health outcomes while reducing healthcare system costs will require increasing emphasis on prevention, personalized medicine and prediction. Drawing on the airline experience, the recent proliferation of wearable devices to continuously monitor physiological parameters offers promise of a technological solution to improve health outcomes and the healthcare system.

In 2013, ASU's Center for Sustainable Health (now the Biodesign Pathfinder Center) launched a research initiative, Project Honeybee, to test the utility of wearable physiological sensors (devices) in the clinical management of patients through a series of small observational clinical trials (OCT). Each Honeybee OCT began with a clinical question and embedded a highly experienced doctoral level Graduate Research Assistant in order to facilitate workflow as well as train the health workforce of the future. We named the effort after the honeybee, which is thought to be nature's best collector of information. Just as honeybees gather data from nature and return collected knowledge to the hive, we sought to investigate the potential of wearable devices to gather information

✉ Jeffrey T. La Belle
JEFFREY.LABELLE@asu.edu

[1] Biodesign Pathfinder Center, Biodesign Institute at Arizona State University, Tempe, AZ 85281, USA

[2] School for the Future of Innovation in Society, Tempe, AZ 85287, USA

[3] College of Nursing and Health Innovation, Tempe, AZ 85004, USA

[4] School of Biological and Health Systems Engineering, Arizona State University, Tempe, AZ 85287, USA

[5] School of Medicine, Mayo Clinic Arizona, Scottsdale, AZ 85259, USA

from a patient's daily living and return it to the clinical "hive" to better inform the healthcare experience.

Funded by ASU for a series of four years, Project Honeybee OCTs incorporated ACO quality measures and performance standards in each of our yearly trials as a value-added metric and service. A key feature of our effort has been the interdisciplinary support from ASU's innovative nursing, engineering, business, and design leaders engaging in research.

We had several goals for these trials, including expanding ASU engagement with clinicians in the greater Phoenix area, initiating collaborations between ASU faculty and area clinical institutions, and testing the challenges that implementing sensors in clinical medicine would face. We hoped to gain insights into the functioning of commercial devices, the needs of clinical teams implementing devices, patient compliance with device monitoring, physician engagement, the value of data obtained from wearable devices, and the range of applications devices could be utilized to monitor.

2 Methods

2.1 Clinical trials

Each Spring we issued a Call for Proposals (CFP) to ASU investigators and area clinicians. Our CFP specified that we supported observational clinical trials of wearable devices available at a cost of 50–300 USD with anticipated study enrollment of 20–50 participants during the 1-year funded period of each study. The CFP clarified that we would not support interventional trials for devices that did not have FDA clearance. The CFP further specified the requirement for qualified clinician leadership by a healthcare provider in Maricopa county. The expected outcome for each study was pilot data suitable for publication and/or further external funding applications.

We sought to maximize experiences while minimizing costs. Each trial was allotted 27,000 USD to support a doctoral nursing student half-time for one year to manage patient enrollment and data collection, and 3000 USD for procuring devices. We did not support Principal Investigator (PI) salaries, clinical time, or patient incentives. Consequently, each trial was limited both in the number of patients and the duration of the trial. However, the many short trials approach provided very diverse experiences across the range of clinical conditions monitored.

In spite of the limited support, our call for proposals was nearly fully subscribed each year. The range of devices ad study approaches was broad. In addition to observational clinical trials, the advisory committee elected to fund some studies each year that did not precisely conform to the CFP requirements for a clinical observational trial including some pre-clinical validation studies, algorithm development studies, and novel device development. Tables 1, 2, 3, and 4 below provide an overview of the funded projects for the four years of Project Honeybee.

2.2 Database

In addition to the clinical studies, Project Honeybee also funded the development of a database of commercially available devices. The database was constructed by conducting regular searches of consumer and trade publications, as well as weekly Google searches to identify consumer wearables in development and/or available for purchase. The database captured device name, manufacturer information, cost, published technical specifications, and target biosignals. The intention was for the database to be a searchable resource for clinicians to identify options for wearable monitors to recommend to patients.

We employed one graduate engineering student for 20 h per week for database construction and maintenance. Over the period of two years, we found that rapid changes in the field of commercial wearables exceeded the work capacity of a single 0.5 fulltime equivalent (FTE) worker and would require at least 1 fulltime staff to manage the constant changes. We explored the possibility of commoditizing the database to support the resources necessary to fund the development and ongoing maintenance of a sophisticated and clinically-relevant searchable platform. After several months of early business model development, we determined that the ongoing database project did not fit with the long-term goals of Project Honeybee at this time, and we discontinued the database project.

3 Results

As Project HoneyBee consisted of 32 unique studies, for the sake of brevity we will highlight two very different ones. An early study on the evaluation of the wearables to support the device database to aid the other studies will be highlighted as well as one of the clinical evaluations of placing devices on patients.

3.1 Project honeybee preclinical trial

3.1.1 Goals

We sought to evaluate the performance of commercially available wearable devices against a gold standard in a laboratory-based rest and exercise protocol. In addition, we evaluated each device with innovation, quality, cost, and comfort scores derived by our research team.

Table 1 Observational clinical trials 2014–2015

Funded Trials - Year 1

Project	Principal investigator(s)	Clinical partner(s)	Device(s)	Endpoints	Subjects enrolled
Testing patient mobility as a predictor of 30-day hospital readmission of heart failure patients	Dr. Matthew Buman GRA: Erin Krzywicki, RN, BSN	Scottsdale Healthcare John C. Lincoln Dignity Health	activPAL3c (PAL Technologies, Ltd., Glasgow, Scotland, UK) sensor (in-hospital), Tractivity (Kineteks Corp., Warwick, RI, USA) monitor (in-hospital), Jawbone (Jawbone, San Francisco, CA, USA) activity monitor (post-hospital)	Correlate inpatient mobility to post-discharge activity	29
Testing arrhythmia self-monitoring of atrial fibrillation to reduce hospitalization	Dr. Heather Ross GRA: Alyssa Smith, RN, BSN	Arizona Arrhythmia Consultants	AliveCor monitor (AliveCor, Inc., Mountain View, CA, USA	Self-assessment accuracy of heart rhythm	18
Validating self-monitoring devices against intrathoracic impedance measurements from implantable cardiac devices for self-monitoring of heart failure	Dr. Heather Ross GRA: Lexy Cook, RN, BSN	Arizona Arrhythmia Consultants	Wello monitor (Azoi, Inc., San Francisco, CA, USA) [note the Wello monitor was never shipped] Jawbone Up3 (Jawbone, San Francisco, CA, USA)	Validation of resting heart rate Correlation of resting heart rate data to changes in fluid status	4
Long-term gait monitoring to improve patient selection for shunt surgery in normal pressure hydrocephalus	Dr. Narayanan Krishnamurthi Dr. James Abbas et al. GRA: Debra Corley, RN, BSN	St. Joseph's Hospital & Medical Center Barrow Neurological Institute	Inertial Measurement Units (IMUs) (proprietary device, Krishnamurthi Lab, Phoenix, AZ, USA)	Gait assessment	0
Heart Failure	Dr. Steve Lester GRA: Mary Kourieh, RN, BSN	Mayo Clinic	Withings Smart Scale, Blood Pressure Cuff and Activity Tracker (Withings, Issy-les-Moulineaux, Île-de-France, France)	Quality of life, Self-efficacy	Not available
Feasibility study for physical activity monitoring devices for diabetes patients	Dr. Jim Levine Dr. Curtiss Cook GRA: EJ Milas, RN, BSN	Mayo Clinic	KAM (Kersh Health, Richardson, TX, USA)	Acquisition of physical activity data	17
Daily energy and neurocognitive function	Dr. Jim Levine	Mayo Clinic	Apple Watch (Apple, Inc., Cupertino, CA, USA)	Activity	0
Pre-clinical trial to validate patch and wearable devices against state of the art (SOTA) seeking gold standard.	Dr. Jeffrey La Belle Dr. Mark Spano GRA: Kevin Uchimura	Mayo Clinic	Multiple patch and ambulatory devices: • Braun forehead thermometer (Braun, Kronberg, Germany) • Braun thermoscan thermometer (Braun, Kronberg, Germany) • Mabis digital thermometer (Briggs Healthcare, Waukegan, IL, USA) • Omron 5 series (Omron Healthcare, Kyoto, Japan) • Omron 7 series (Omron Healthcare, Kyoto, Japan) • Omron HEM-18 (Omron Healthcare, Kyoto, Japan) • KAM (Kersh Health, Richardson, TX, USA) • Withings wireless BP monitor (Withings, Issy-les-Moulineaux, Île-de-France, France) • Fitbit One (Fitbit, San Francisco, CA, USA) • AliveCor (AliveCor, Inc., Mountain View, CA, USA) • Tanita BF 679 W (Tanita Corporation, Tokyo, Japan) • VitalConnect Healthpatch Biosensor (VitalConnect, San Jose, CA, USA) • Fitbit Surge (Fitbit, San Francisco, CA, USA) • Withings body composition monitor (Withings, Issy-les-Moulineaux, Île-de-France, France) • Tinke (Zensorium, Singapore) • Finger Pulse OX SM-110 (Santamedical, Tustin, CA, USA) • Withings Pulse O2 (Withings, Issy-les-Moulineaux, Île-de-France, France) • FeverSmart (Fever Smart, Philadelphia, PA, USA) • GENEActiv Original (Activinsights, Kimbolton, Cambridgeshire, UK) • Jawbone Up3 (Jawbone, San Francisco, CA, USA) • Apple Watch (Apple, Cupertino, CA, USA)	Lab validation	38 participants, generating ≥20 data sets per device

Table 2 Observational clinical trials 2015–6

Funded Trials - Year 2

Project	Principal investigator(s)	Clinical partner(s)	Device	Endpoints	Subjects enrolled
Develop a machine-learning based algorithm to quantify type, context, and intensity of wearable device data collected at wrist and hip. Make algorithm available to other PHB teams and the research community at large. ACO-5, ACO-7	Matthew Buman, PhD Pavan Turaga, PhD CO-PI: Brian Ashcroft, PhD	ASU, SNHP ASU, SECEE ASU, Biodesign, CSMB	GENEactiv Actigraph GT3x+ (Activinsights, Kimbolton, Cambridgeshire, UK)	Machine learning algorithm development	0, secondary data analysis only)
Measure the impact of outpatient Red Blood Cell (RBC) transfusions on at-home functional status through physiological parameters and physical activity metrics. ACO-7	Daryl Kor, MD Jyotishman Pathak, MD CO-PI: Karl Poterack, MD	Mayo Clinic Rochester Mayo Clinic Rochester Mayo Clinic Scottsdale	Fitbit Charge HR (Fitbit, San Francisco, CA, USA)	Post-op activity assessment	4
Long-term continuous monitoring to characterize the effects of deep brain stimulation on gait and balance dysfunction in people with Parkinson's disease ACO-6, ACO-7	Narayanan Krishnamurthi, PhD CO-PI: James Abbas, PhD PARTNERS: Padma Mahant, MD Francisco Ponce, MD	ASU, CONHI ASU, SBHSE Banner Good Samaritan Medical Center St. Joseph's Hospital and Medical Center	Inertial Measurement Units (IMU) (proprietary device, Krishnamurthi lab, Phoenix, AZ, USA)	Gait and balance assessment	3
Verification and validation of customized device sensors to reduce hospital readmission to demonstrate equivalency between custom array and gold standard in the lab. ACO-1, ACO-2, ACO-5, ACO-6, ACO-7	Jeffrey La Belle, PhD Mark Spano, PhD CO-PI: Eric Steidley, MD	ASU, SBHSE ASU, SBHSE Mayo Arizona	Customized lab-built sensor array/ device monitoring activity, temperature, ECG and pulse oximetry (La Belle lab, Tempe, AZ, USA)	Sensor validation	10
To create a scientifically validated weight loss program to reach patients utilizing wearable technologies. ACO-5, ACO-7	James Levine, MD CO-PI: Gabriel Koepp	Mayo Clinic Arizona / ASU Apple, Inc. SP Health	Apple Watch (Apple, Cupertino, CA, USA)	Activity, Weight change	0
Defining the healing trajectory after elective surgery. Use wearable activity monitoring devices pre and post elective inguinal hernia surgery (2 types) to ascertain quality/duration of return to baseline activity and identify patients at risk for opioid dependence. ACO-2, ACO-5, ACO-6, ACO-7	Karl Poterack, MD CO-PI: Richard T. Schlinkert, MD CO-PI: Matthew Buman, PhD	Mayo Clinic Arizona Mayo Clinic Arizona ASU	GENEactiv (Activinsights, Kimbolton, Cambridgeshire, UK)	Activity	7
Assess the feasibility and effectiveness of patient-directed deployment of an individualized medication management plan to treat recurrences of Atrial Fibrillation (AF) utilizing wearable monitor. ACO-1, ACO-4, ACO-5, ACO-6, ACO-7	Heather Ross, DNP Partners: Mitchell J. Ross, MD John Beshai, MD	Arizona Arrhythmia Consultants / ASU Arizona Cardiology Group Mayo Clinic Arizona	AliveCor ECG monitor (AliveCor, Mountain View, CA, USA)	Quality of life, Healthcare utilization	8
Assess the feasibility and effectiveness of patient-directed Congestive Heart Failure (CHF) assessment and management using an ambulatory health tracking device. ACO-1, ACO-4, ACO-5, ACO-6, ACO-7	Heather Ross, DNP Partners: Mitchell J. Ross, MD Eric Steidley, MD John Beshai, MD Mitchell J. Cohen, MD Edward Rhee, MD	Arizona Arrhythmia Consultants / ASU Arizona Cardiology Group Mayo Clinic Arizona Mayo Clinic Arizona Phoenix Children's Hospital Phoenix Children's Hospital	Jawbone Up3 (Jawbone, San Francisco, CA, USA)	Resting heart rate validation	11
Early recognition of sepsis to reduce mortality, organ failure, hospital length of stay, and cost of care. Monitor relevant physiological trends with high signal-noise ratio and determine wearable device feasibility of use in a high-risk Hematopoietic stem cell transplantation (HSCT) cohort. ACO-1, ACO-7	Ayan Sen, MD Pierre Noel, MD Charles Bruce, MD CO-PIs: Michelle Alore, RN, MSN, DNP Bhavesh Patel, MD Jane Olsen, NP Brittany Williams, RN	Mayo Clinic Arizona Mayo Clinic Arizona Mayo Clinic Arizona Mayo Clinic Arizona Mayo Clinic Arizona Mayo Clinic Arizona	Body Guardian (Preventice Technologies, Rochester, MN, USA)	Heart rate validation	11

Table 3 Observational clinical trials 2016–7

Funded Trials - Year 3

Project	Principal investigator(s)	Clinical partner(s)	Device	Endpoints	Subjects enrolled
Assessing the feasibility and effectiveness of using a wearable health tracking device to identify Military Veterans at risk for service-related conditions	Dr. Tamiko Azuma Karen Gallagher, MS Dimple Patel GRA	ASU Pat Tillman Center ASU Speech and Language Clinic	Vivofit HR (Garmin, Olathe, KS, USA)	Correlation of activity and sleep with self-rated stress, anxiety, sleep, and cognition	21
Toward a portable wearable system for detection and diagnosis of obstructive sleep apnea	Dr. Matt Buman Dr. Pavan Turaga Maynak Gupta GRA	None	Zephyr BioHarness (Medtronic, Minneapolis, MN, USA) GeneActiv (Activinsights, Kimbolton, Cambridgeshire, UK)	Algorithm development	0 (secondary data analysis only)
Mayo 360 Personalized Support System (Mayo 360PSP). Patient-Driven Lifestyle Solutions for Weight and Wellness.	Amy Foxx-Orenstein, DO Dr. George Runger Tanya Carroccio GRA	Mayo Clinic Arizona Weight and Wellness Solutions Program	Fitbit Charge HR (Fitbit, San Francisco, CA, USA)	Activity measurement	27
Monitoring to Investigate the Effects of Sleep Behavior Disorder on Gait Dysfunction in People with Parkinson's Disease	Dr. Narayanan Krishnamurthi Stephanie De Santiago GRA	St. Joseph's Hospital and Medical Center	Actiwatch Spectrum (Koninklijke Philips N.V., Amsterdam, The Netherlands) APDM Inertial Measurement Units (proprietary device, Krishnamurthi lab, Phoenix, AZ, USA)	Sleep amount and gait correlation	0
Expand development on ambulatory medical wearable device captures three biosignals for patients	Dr. Jeffrey La Belle Heather Ross, DNP Alejo Chavez/Ashwin Sundar GRA's	None	Design and fabricate a device in-house to measure outcomes	Device fabrication	3
Measurement of medical staff and trainees' sleep and physical activity during various call and shift schedules	Dr. Karl Poterack Dr. Andrew Gorlin Dr. Matt Buman Ran McBryde GRA	Anesthesiology Department at Mayo Clinic in Phoenix, AZ	GENEactiv (Activinsights, Kimbolton, Cambridgeshire, UK)	Sleep amount and quality	11
ACT NOW: Use of ACT9X activity and sleep tracker to help with early detection of Alzheimer's disease in Older adults with Down syndrome	Dr. Shannon Ringenbach Nathaniel Arnold GRA	Alzheimer's and Cognitive Disorders Program, Barrow Neurological Institute	Actigraph GT9X Link (Actigraph LLC, Pensacola, FL, USA)	Physical activity and sleep correlation Determination of effects of activity training on physical activity and sleep	15
Validating Force-Sensing Mouthguards Against Expert Concussion Analysis to Guide Removal from Play in Collegiate Athletics	Heather Ross, DNP Lexy Richards GRA Alyssa Smith GRA	ASU Athletics	FITGuard mouthguard (Force Impact Technologies, Los Angeles, CA, USA)	Head trauma detection	5
Wearable wireless vital signs monitoring in predicting clinical decompensation	Dr. Ayan Sen Dr. Holly Geyer Dr. Brian Hull Dr. Christopher Farmer Madeleine Goetz GRA	Mayo Clinic Arizona	Zephyr BioHarness (Medtronic, Minneapolis, MN, USA)	Heart rate, Respiratory rate, Heart rate variability, Clinical decompensation	11

Table 4 Observational clinical trials 2017–8

Funded Trials - Year 4

Project	Principal investigator(s)	Clinical partner(s)	Device	Endpoints	Subjects enrolled
Patient-Driven Lifestyle Solutions for Weight and Wellness – Part 2	Amy Foxx-Orenstein D.O. George Runger, PhD, Julie Liss, PhD, Matthew Buman, PhD (Co-PIs) Tanya Carroccio GRA Miriam Zerio-Levering	Ken and Linda Morris Weight and Wellness Solutions Program at Mayo Clinic, Arizona	Fitbit Charge HR (Fitbit, San Francisco, CA, USA)	Activity measurement	27 (continued from Year 3)
Patient-Driven Lifestyle Solutions for Weight and Wellness – Development of a Software Platform	Amy Foxx-Orenstein D.O. George Runger, PhD, Julie Liss, PhD, Matthew Buman, PhD (Co-PIs) Tanya Carroccio GRA Miriam Zerio-Levering	Ken and Linda Morris Weight and Wellness Solutions Program at Mayo Clinic, Arizona	Data from the Fitbit Charge HR (Fitbit, San Francisco, CA, USA) will be utilized in the creation of an online dashboard	Dashboard software development	0 (secondary data analysis)
Feasibility and Explorative Study to test the superiority of Connected Health Care using wearable devices to improve body composition and fitness levels	Regis Fernandes, MD Robert Scales, MS, Minako Katayama, MD, Hari Chaliki, MD, Sophie Bersoux, MD, Temkit Hamy, PhD (CO-PIs) GRA - TBD	Robert Scales, MS Hari Chaliki, MD Sophie Bersoux, MD	24alife (Mikropis Holding, Zalec, Slovenia) FitBit Zip (Fitbit, San Francisco, CA, USA) Heart-rate chest sensor (TBD)	Activity, Weight Change	17
Metabolic Tracking During Pregnancy to Promote Adequate Gestational Weight Gain	Elizabeth Reifsnider, PhD, RN Corrie Whisner, PhD, Kiley Vander Wyst, MPH (Co-PIs) Kiley Vander Wyst - GRA	Adelante Healthcare	Breezing Device (Breezing Metabolism, Tempe, AZ, USA)	Pregnancy weight gain	23
Real-World Validation of a Novel Multi-sensor Wearable for Heart Failure	Jeffrey La Belle, PhD Matthew Buman, PhD Heather Ross, PhD, DNP Pavan Turaga, PhD (Co-PI) Raja Nagappan - GRA Alexan Gomez - GRA Robin Faulkner - GRA	Honor Health Heart & Vascular Institute	Design and fabricate a multi-sensor CHF device in-house	Device validation	0
Validation of Force and Vector in Athletic Head Contact	Heather Ross, PhD, DNP Lexy Richards, RN, Carter Pruitt, Joshua Beaumont (Co-PIs) Austin Tielke - GRA Dimple Patel - GRA	ASU Athletics ASU Wrestling	IM Mouthguard (Prevent Biometrics, Edina, MN, USA)	Head trauma validation	4

3.1.2 Methods

We recruited participants from the ASU student community to perform multiple "typical use" protocols with 22 different devices across 7 categories (Table 5) with 15 unique sensor types (Table 6). Table 5 illustrates the total data collection that took place during the 12-month funding period. We achieved our goal to collect a minimum 20 basal and post-exercise readings (data sets) from unique participants for all devices when possible. Only two devices were tested less frequently: the VitalConnect Healthpatch Biosensor, a device now not available to the consumer ($n = 9$), and the FeverSmart wearable thermometer (n = 9), which had various issues during use, preventing accurate data collection.

Testing Protocols:

Activity Testing: During activity testing, the participant in the study was filmed from the waist down for the duration of the protocol. The participant was instructed to walk naturally at five different speeds. Then, they walked at each speed for one minute before changing to the next. The speeds were 2.5, 3.0, 3.4, and 4.0 mph. The participant wore several devices on their wrists, including the GENEActiv, Apple Watch, Jawbone UP3, and Fitbit Surge. The Fitbit One, KAM, and Withings Pulse O2 were worn on a fitness belt that included a pocket for the iPhone 6 (necessary for use with the Apple Watch). The total steps collected by each device was recorded before and after exercise; total steps walked during the five-minute session were compared to an average manually counted by three investigators separately.

Blood Pressure Testing: Blood pressure testing began with participant seated, at rest while the BPMs were used. First, the participant's manual blood pressure was taken on their right or left arm. The rest of the devices were randomized before testing began to ensure no bias in time effect after exercise. According to specific device use, the Omron 7 was always used on the left wrist and the other devices, Omron 5 and Withings Wireless, were used in succession on alternating arms. These devices were the Omron 5 and the Withings Wireless. Once each device was tested, the participant was asked to run for five minutes on a treadmill, beginning at a speed of 2 mph and increasing or decreasing at the participant's preference. Immediately after running, the devices were tested in the same order the specific participant used during baseline testing. If this protocol followed in succession from a different protocol, the participant rested for 10 min to ensure baseline reading accuracy. Time stamps were recorded for every measurement taken.

GSR/Total Body Fat Testing: While the participant was at rest, an investigator applied iWorx GSR leads to their index and middle finger and information was collected from the participant to ensure the settings of the devices were personalized. With shoes and socks removed, data were continuously recorded using the iWorx. The participant first was instructed to use the Tanita scale followed by the Withings Smart Body Analyzer. Each scale was used three times, each in alternating succession. With their shoes back on, the participant was asked to run for five minutes on a treadmill, beginning at a speed of 2 mph and increasing or decreasing at the

Table 5 Devices tested in preclinical trial

Devices to be evaluated	Data sets collected	Device usage category
Braun Forehead Thermometer	32/20	Thermometer
Braun Thermoscan Thermometer	32/20	Thermometer
Mabis Digital Thermometer	32/20	Thermometer
Omron 5 series	31.5/20	Blood Pressure Monitor
Omron 7 series	31.5/20	Blood Pressure Monitor
Omron HEM-18	30/20	Blood Pressure Monitor
KAM	34/20	Activity Monitor
Withings Wireless BP Monitor	23/20	Blood Pressure Monitor
Fitbit One	28/20	Activity Monitor
AliveCor	22/20	EKG/Heart Rate Monitor
Tanita BF 679 W	23/20	Bioelectrical Impedance Analysis
VitalConnect Healthpatch Biosensor*	9/9	Thermometer, Activity Monitor, Respiratory Rate Monitor
Fitbit Surge	29/20	Activity Monitor, Heart Rate Monitor
Withings Body Composition Monitor	22/20	Heart Rate Monitor, Bioelectrical Impedance Analysis
Tinke	22/20	Pulse Oximeter, Respiratory Rate Monitor
Finger Pulse OX SM-110	21/20	Pulse Oximeter
Withings Pulse O2	30/20	Activity Monitor
Withings Pulse O2	21/20	Pulse Oximeter
FeverSmart*	9/9	Thermometer
GeneActiv Original, Sleep, Action, Wireless	22/20	Activity Monitor
Jawbone UP3	20/20	Activity Monitor
Apple Watch	20/20	Activity Monitor, Heart Rate Monitor
Braun Forehead Thermometer	32/20	Thermometer

Table 6 Sensor types and references for biosignals of interest

Biosignal	Sensor Type 1	Sensor Type 2	Sensor Type 3	Sensor Type 4	Gold Standard Method/Reference
Temperature	Oral thermometer	Tympanic thermometer	Infrared thermometer	Auxiliary thermometer	Interdevice comparison
Activity	3-axis accelerometer	3-axis accelerometer, 3-axis gyroscope, & compass	N/A	N/A	Video movement observation
Blood Pressure	Inflatable arm cuff	Inflatable wrist cuff	Wireless inflatable arm cuff	N/A	Manual blood pressure
Respiratory Rate	3-axis accelerometer	LED	N/A	N/A	Research grade spirometers
EKG	Single-lead EKG	N/A	N/A	N/A	Research grade 3-lead EKG and self-comparison
GSR	Single-lead GSR	N/A	N/A	N/A	Research grade 2-lead GSR
SpO_2	Infrared LED	Spectro-photometer & reflective films	Research grade finger pulse oximeter	N/A	Research grade finger pulse oximeter

participants' preference. Immediately after exercise, with shoes and socks quickly removed, the same steps were repeated with continuous iWorx data collection and alternating scale measurements. If this protocol followed in succession from a different protocol, the participant rested for 10 min to ensure baseline reading accuracy. Time stamps were recorded for every measurement taken.

EKG Testing: While the participant was at rest, an investigator applied the iWorx 3-lead EKG system to the participant's wrists and right ankle. Due to possible noise, the participant was instructed to sit at about six feet away from iWorx module and desktop computer. Then, the AliveCor was given to the participant and the iWorx was started. The AliveCor collected a total of three readings, 30 s each, in succession. The AliveCor was used until three successful files were written. The time the readout occurred was recorded for later iWorx processing. When all three readings were completed, the participant was asked to run for five minutes on a treadmill, beginning at a speed of 2 mph and increasing or decreasing at the participant's preference. Immediately after exercise, the same steps were repeated with continuous iWorx data collection and AliveCor measurements. If this protocol followed in succession from a different protocol, the participant rested for 10 min to ensure baseline reading accuracy. Time stamps were recorded for every measurement taken.

Pulse Oximetry Testing: While the participant was at rest, the SM-110 was applied to their right index finger and the iWorx pulse oximeter module on their right middle finger. With the iWorx recording and SM-110 turned on, their left-hand alternated between the Tinke and the Withings Pulse O2. When either the Tinke or Withings Pulse O2 finished a reading, its value and the concurrent SM-110 value were recorded, meaning a total of 3 values for the Tinke and Withings Pulse O2, and a total of 6 values for the SM-110 were recorded. The participant was asked to run for five minutes on a treadmill, beginning at a speed of 2 mph and increasing or decreasing at the participant's preference. Immediately after exercise, the same steps were repeated with continuous iWorx data collection and device measurements. If this protocol followed in succession from a different protocol, the participant rested for 10 min to ensure baseline reading accuracy. Time stamps were recorded for every measurement taken.

Respiratory Rate Testing: The participant was asked to rest for ten minutes before testing began. This allowed time for the breathing rate to normalize and the iWorx spirometer to calibrate. The spirometer was set up on a ring stand for the participant's comfort and to reduce noise from movement. The participant was instructed to wear a nose clip during the data collection periods. After the 10-min rest and initialization period, the participant was asked to breathe entirely through the spirometer while the Tinke was used on their left hand three times in succession. The participant was asked to run for five minutes on a treadmill, beginning at a speed of 2 mph and

increasing or decreasing at the participant's preference. Immediately after exercise, with the nose clip reapplied by the participant, the same steps were repeated with continuous iWorx data collection and Tinke measurements. If this protocol followed in succession from a different protocol, the participant rested for 10 min to ensure baseline reading accuracy. Time stamps were recorded for every measurement taken.

Temperature Testing: While the participant was at rest, the FeverSmart was applied to their underarm with sticky gauze. The Mabis Digital Oral thermometer was first placed in the participant's mouth while an investigator used the Braun Forehead thermometer on the participant. Then, the Braun Thermoscan was handed to the participant for them to insert in their ear. Each device was used three times in the order described previously. The participant was asked to run for five minutes on a treadmill, beginning at a speed of 2 mph and increasing or decreasing at the participant's preference. Immediately after exercise, the devices were used again in the same order. When all three values were recorded for each device, the data from the FeverSmart were collected. If this protocol followed in succession from a different protocol, the participant rested for 10 min to ensure baseline reading accuracy. Time stamps were recorded for every measurement taken.

VitalConnect Testing: For the first part of the VitalConnect Protocol, the VitalConnect patch was applied to their chest, just below the heart in the orientation indicated in the manual. The FeverSmart patch from Temperature Testing was also applied in the same way as before. While the participant was at rest, the iWorx spirometer was set up to calibrate and allowed the participant's breathing rate to normalize. The spirometer was set up on a ring stand for the participant's comfort and to reduce noise from movement. The participant was instructed to wear a nose clip during the data collection periods. After the 10-min rest and initialization period, the participant was asked to breathe entirely through the spirometer while the Tinke was used on their left hand three times in succession. The participant's temperature was also recorded using the Braun Thermoscan and the Braun Forehead thermometer in alternation. The participant was asked to run for five minutes on a treadmill, beginning at a speed of 2 mph and increasing or decreasing at the participant's preference. Immediately after exercise, with the nose clip reapplied by the participant, the same steps were repeated with continuous iWorx data collection, Tinke, and thermometer measurements. If this protocol followed in succession from a different protocol, the participant rested for 10 min to ensure baseline reading accuracy. Time stamps were recorded for every measurement taken.

Immediately following the aforementioned VitalConnect Testing, a second part was completed. During activity testing for the VitalConnect, the VitalConnect patch was still continuously worn, and the steps described previously for Activity Testing were followed.

Innovation Score: Device innovation was rated on a scale of 0 to 5 and represents an average of equally weighted scores that were given to the device that include ease of use, design aesthetics, invasiveness, wireless capabilities, operating constraints, and size. Invasiveness, wireless capabilities, operating constraints, and size had a defined range based on manufacturer metrics provided for each device. Ease of use and design aesthetics were the only categories that were subjective to the study participants, as they were the device operators. Participants were asked a series of questions to determine how intuitive the device was to use, ease of operation, clarity of formal instructions, clarity of the interface and menus, and ease of interpreting the data. Design aesthetics averaged the opinions of all participants who used the device with respect to their idea of the device's overall attractiveness, and how attractive the device is in its class. The specific questions the participants were asked included the following, each with its own 0 to 5 rating.

1. How intuitive is the device to use?
2. How easy is it to hold and successfully and safely operate it at the same time?
3. How clear are the formal instructions?
4. How clear is the device interface or menus?
5. How clearly are the results communicated?
6. How attractive is the device design in general?
7. How attractive is the device compared to other devices in its class?

Comfort Score: Like the innovation score, comfort was quantitatively evaluated by our participants. After a participant completed a test, they were given a form that includes the following scale. The participant was able to view the scale before testing so that they could think about each device in the following terms:

9. Unnoticeable
8. Occasionally Noticeable
7. Constantly Noticeable
6. Occasionally Annoying
5. Constantly Annoying
4. itchy Irritant
3. Concerning Pressure
2. Minor Pain
1. Major Pain

This score of 0 to 9 was then converted to a scale of 0 to 5 and weighted equally in the invasiveness category within the innovation score.

Quality Score: Device quality was rated on a scale of 0 to 5 and represents an average of scores that were given to the device that include data accessibility and performance. Data accessibility represents the average value assigned for

accessibility to raw data, memory of results, and exportability of data. Performance was designated as a function of reliability and reusability, which was again assigned a score of 0 to 5.

Reliability scores were based on the p values for hypotheses tested, and the slope and the corresponding R^2 of the gold standard graphs (GSG). Reusability was defined as how quickly the device could repeat a measurement as well as the relative standard deviation (RSD) found in baseline testing where applicable.

Cost Score: The cost score for each device was updated based on market prices for the devices on 08/13/2015. The actual cost score relates the cost of the device overall as well as its relative cost in its class wherein the overall cost is weighted more heavily than the relative cost. In this way, the most important aspect is general cost, but the cost of similar devices is still considered.

3.1.3 Data analysis

Basal variance and relative standard deviation For all devices where replicate measurements were taken at rest (basal readings of at least $n = 3$), we determined population averages, standard deviation, and RSD. RSD is a comparison of the average and standard deviation that measures how precise the average is and measures random error.

Gold standard graphical comparisons Similar to the Yellow Springs Instrument for blood glucose, clinical GS methods exist for other physiological measurements. With respect to the 7 biosignals of interest to this OCT (Table 6), both GS devices and reference methods were used to evaluate commercial device performance. Manual blood pressure, core body thermometers, multi-lead EKG and galvanic skin response (GSR) systems, finger pulse oximeters as well as blood gas analyzers, and spirometers, are widely used as clinical GSs for assessment in the hospital. The reference used for quantifying physical activity was direct observation and monitoring. A simple test system, using versions of the clinical GS devices listed above, allowed for gold standard graphs (GSG) to be created for each device tested.

Only three devices were not able to be calibrated directly against a true GS (i.e. AliveCor ECG monitor, Tanita scale [using GSR for calculating body fat percentage (BF%) and water percentages], and Withings Smart Body Analyzer [using GSR for calculating BF% and water percentages].

In the case of the AliveCor, raw signal data were not available. The GS method, a three lead EKG system, was taken alongside the AliveCor measurements, and side-by-side images were evaluated for the clear detection and counting of QRS complexes (represented by tick marks) and associated waves. In order to evaluate the device's accuracy in the presence of noise or other interference, the total tick marks for the 30 s period are counted and recorded. The total number of QRS complexes is also counted within the same document. Any QRS complex that does not have a corresponding tick mark is recorded as well as any tick mark that does not have a corresponding QRS complex. These particular tick marks are recorded as *miscounted ticks* and this method was used as the reference for comparison of EKG.

For Tanita and Withings scales, the GSR recordings collected via the iWorx unit did not provide the expected data. The data collected from the iWorx was categorized as the skin impedance change in relation to emotional stress. We found that physical stress (as a result of exercise via the scales) and emotional stress (via iWorx) data are not directly comparable. As a result, we developed an alternative reference method to use with the Withings and Tanita body fat percentage readings. The participant's demographic information was used to predict a body fat percentage using information gained from their pre-testing survey and equations below, taken from "Body mass index as a measure of body fatness: age-and-sex specific prediction formulas" (Deurenberg et al., 1991):

$$BMI = \frac{weight(kg)}{height(cm)^2} \tag{1}$$

$$(1.2^{*}BMI) + (0.23^{*}age) - (10.8^{*}gender) - 5.4$$
$$wherein, 1\ is\ coded\ for\ "male"\ and\ 0\ is\ coded\ for\ "female" \tag{2}$$

GSGs were created using the references listed above, and error zones were created by estimating between 5% and 20% error at the lowest and highest possible physiological levels This standard can determine the device's relationship to its reference thereby elucidating the dynamic range, percent data off the calibration line, upper and lower limits of detection, accuracy, and precision of the device.

Testing for statistical significance In order to measure the significance of exercise, t-tests were conducted on all devices for their readings before and after exercise. We analyzed data from all devices to test the hypothesis that exercise will cause a significant change in blood pressure ($p < .05$ indicates statistical significance). A second set of t-tests were conducted to test the hypothesis that a specific device differs significantly from its reference method ($p < 0.05$ indicates statistical significance). The associated p values for each device with respect to each hypothesis and reference used can be seen in Table 7.

3.1.4 Results

Best-in-class Using the innovation, cost, and comfort scores as seen in Table 7, the best-in-class devices were identified in each category: Braun Thermoscan (temperature), Apple Watch (activity), Omron 5 arm meter (blood pressure), VitalConnect Healthpatch Biosensor (respiratory rate),

Table 7 Summary of statistical results and analysis for all devices

Name	Category	Innovation	Quality	Cost	Comfort	Device Name	Baseline v. Reference	Exercise v. Reference	Baseline v. Exercise	N
Braun Forehead †	Thermometer	3	2	4	8	Mabis Digital Temp	5.64E-04	0.19	1.71E-07	32
Braun Thermoscan †	Thermometer	2	2	3	8	Mabis Digital Temp	2.79E-05	3.46E-05	0.16	32
Mabis Digital †	Thermometer	2	1	5	6	Braun Thermoscan	2.79E-05	3.46E-05	0.06	32
FeverSmart †	Thermometer	4	3	1	8	Mabis Digital Temp	0.23	0.19	0.57	8
Vital Connect Healthpatch Biosensor †	Thermometer	4	4	0	8	Braun Thermoscan	1.39E-06	3.21E-06	0.45	8
Withings Wireless BP Monitor ‡	Blood Pressure Monitor	4	3	1	6	Omron HEM-18	0.33/0.69	6.16E-05/0.19	0.02/0.005	21
Omron 5 series	Blood Pressure Monitor	3	3	4	7	Omron HEM-18	0.01/0.62	0.31/0.13	0.0004/0.06	30
Omron 7 series ‡	Blood Pressure Monitor	3	3	3	7	Omron HEM-18	0.18/0.16	0.003/0.73	0.002/0.08	30
KAM ¶	Activity Monitor	2	3	4	9	Steps from Manual step count	–		–	38
Fitibit One	Activity Monitor	4	3	3	9	Steps from Manual step count	–	0.34	–	28
Withings Pulse O2 †	Activity Monitor	4	3	4	9	Steps from Manual step count	–	0.01	–	33
Fitbit Surge †	Activity Monitor	4	3	0	8	Steps from Manual step count	–	0.002	–	30
Vital Connect Healthpatch Biosensor	Activity Monitor	4	4	0	8	Steps from Manual step count	–	0.002	–	8
Alivecor	EKG Monitor	3	4	3	7	iWorx.	0.001	0.001	7.83E-06	21
Tanita BF 679 W †	BIA Monitor	2	3	4	8	Calculated Body Fat	0.002	0.003	0.04	22
Withings Smart Body Analyzer †	BIA Monitor	4	3	1	8	Calculated Body Fat	0.33	0.33	0.58	21
Tinke *	Pulse Oximeter	3	1	2	8	iWorx RR SpO2	2.28E-08	7.60E-07	0.88	20
Finger Pulse OX SM- 110 *	Pulse Oximeter	3	2	5	8	iWorx RR SpO2	6.03E-37	1.70E-40	0.78	20
Withings Pulse O2 *	Pulse Oximeter	4	3	4	8	iWorx RR SpO2	2.60E-05	1.36E-05	0.88	20
Vital Connect Healthpatch Biosensor ¶	Respiration Monitor	4	4	0	8	iWorx RR (BPM)	0.99	0.07	4.05E-05	8
Tinke	Respiration Monitor	3	2	2	8	iWorx RR (BPM)	0.001	5.99E-12	0.021	20
Jawbone UP3	Activity Monitor	4	3	1	8	Steps from Manual step count	–	0.001	–	25
Apple Watch	Activity Monitor	4	3	0	8	Steps from Manual step count	–	0.27	–	25

Device labels		Gold standard graphs						Error		
Name	Category	All Data Slope	All Data R2	N	Device Error Slope	Device Error R2	N	Total Number of Successful Collections	Number of device failures or user errors (data collected)	Number of device or user failures (data not collected)
Braun Forehead †	Thermometer	0.10	0.05	32	0.10	0.02	32	191	0	1
Braun Thermoscan †	Thermometer	0.09	0.04	32	0.24	0.12	32	192	0	0
Mabis Digital †	Thermometer	0.83	0.12	32	0.48	0.12	32	192	19	2
FeverSmart †	Thermometer	0.58	0.08	8	0.58	0.08	8	48	30	4
Vital Connect Healthpatch Biosensor †	Thermometer	0.06	0.02	8	0.06	0.02	8	138	0	0
Withings Wireless BP Monitor ‡	Blood Pressure Monitor	0.63	0.75	21	0.68	0.78	19	42	0	5
Omron 5 series	Blood Pressure Monitor	0.82	0.76	30	0.87	0.77	28	61	0	2
Omron 7 series ‡	Blood Pressure Monitor	0.64	0.65	30	0.68	0.67	28	63	0	5
KAM ¶	Activity Monitor	0.04	0.35	38	0.04	0.35	38	38	0	2
Fitibit One	Activity Monitor	2.51	0.13	30	2.51	0.13	28	30	2	1
Withings Pulse O2 †	Activity Monitor	0.78	0.01	33	0.78	0.01	33	33	1	0
Fitbit Surge †	Activity Monitor	1.4	0.61	30	−0.18	0.01	30	30	0	2
	Activity Monitor	0.78	0.01	8	0.78	0.01	8	8	0	0

Table 7 (continued)

Vital Connect Healthpatch Biosensor						21	126	8		3
Alivecor	EKG Monitor	0.94	0.90	21	0.94	0.90	22	132	1	1
Tanita BF 679 W †	BIA Monitor	1.06	0.31	22	1.08	0.70	21	114	1	5
Withings Smart Body Analyzer †	BIA Monitor	0.98	0.52	21	0.91	0.35				
Tinke*	Pulse Oximeter	−0.09	0.002	21	−0.12	0.003	20	123	1	6
Finger Pulse OX SM-110 *	Pulse Oximeter	0.42	0.44	21	0.42	0.44	20	246	1	0
Withings Pulse O2 *	Pulse Oximeter	−0.09	0.004	21	−0.11	0.01	20	110	1	7
Vital Connect Healthpatch Biosensor ¶	Respiration Monitor	0.87	0.54	8	0.87	0.54	8	8	0	0
Tinke	Respiration Monitor	0.12	0.14	21	0.12	0.16	20	123	1	3
Jawbone UP3	Activity Monitor	0.16	0.01	25	0.34	0.10	25	25	0	1
Apple Watch	Activity Monitor	1.37	0.66	25	1.19	0.73	25	25	0	0

* Quantization in data caused speculation in reliability of slope and R^2 values for this device

† Best fit line given by data shown to be highly speculative

‡ Unknown possible fitting issues seen in gold standard graphs

¶ Major outliers caused fitting issues with best fit line

AliveCor (heart rate/ECG), Tanita BF 679 W (GSR), and SM-110 (pulse oximetry).

3.1.5 Device failures and usability issues

Activity devices The most common failure associated with activity devices was that the devices were not charged, or the devices were not activated by the investigator for collection, which largely falls under user error. With respect to actual device failures, there were three possible errors related to dramatic step undercounting, though it was not clear in some instances if it was user error or actual device malfunction.

Blood pressure devices The BPMs accrued the most failures of all the device types, most often due to user error or user body habitus (i.e. cuff to small for user's arm). The most significant issue that was observed during blood pressure testing was impossibly low or abnormal manual blood pressure readings. In these instances, all of the other BPMs recorded significantly higher blood pressure for the individual participant, and likely represents user error with the manual blood pressure measurement. This observation highlights a challenge in providing gold standard comparison testing for a physiological measurement in a lab where staff may not be trained healthcare providers.

GSR/Total body fat devices GSR devices had errors related to incorrect participant settings resulting in invalid findings. These errors were device-related in failure to connect to the internet, as well as user-related in failure to manually input correct settings.

EKG devices Failure associated with the AliveCor device mainly stemmed from user error, as it was not always easy for the participant to position their hands in the proper orientation and apply the correct amount of pressure to the electrodes. We also observed noise from external conditions that interfered with the high frequency waves that the AliveCor used to transmit the EKG signal.

Pulse oximetry devices We observed numerous device failures with the Tinke, with multiple software crashes and failure to connect to the mobile device. We observed user failures with the Withings Pulse O2, with a participant squeezing the device too firmly.

Respiratory rate devices As the Tinke was used for pulse oximetry testing as well as respiratory rate testing, the software issues associated with this device are listed previously.

Temperature devices We experienced numerous failures with the FeverSmart device, notably related to early battery depletion. Even though the device would go through proper power

down procedures, it would exhaust between 50% and 60% of its battery per test. Another failure with the FeverSmart was due to the sweat buildup on the participant. The adhesive gauze lost its ability to stick to the participant and would commonly lift or fall off completely. These were device failures rather than user error.

The Mabis Digital Oral thermometer demonstrated user error when the participant would open their mouth and talk or not close their moth completely while using the oral thermometer, resulting in impossible hypothermic values. We also observed device error with the Mabis device from an auto-shutoff during testing.

3.1.6 Study summary

Thirty-eight subjects were tested providing a total of 576 data sets that represent 7 different device categories and 22 unique devices. This study provided many opportunities for engineering student investigators to increase their experience in their field, but also for professional development. By working with new technologies, the investigators gain valuable hands-on experience with and exposure to successful wearable devices that can be translated into robust medical devices. Additionally, by working with human subjects the investigators are able to become comfortable when working with a study participant, to understand the need for professional behavior, and to better comprehend the needs of the participant as well as observing for user error. Student investigators also sharpened their critical thinking and problem-solving skills through their thorough analysis of the study data.

This study led to the development of an integrated wireless on-body sensor currently being investigated by the La Belle Labs. Results of this study with best-in-class identifications also provided guidance to other studies related to Project Honey Bee in terms of wearable device selection.

3.2 Heart failure monitoring study

3.2.1 Goals

Building on the notion of lab-based wearable device validation studies, we sought to perform field-based validation of consumer wearable devices using implanted FDA-cleared medical devices (i.e. pacemakers, implantable defibrillators) as the gold standard comparator. We focused on congestive heart failure, using changing cardiopulmonary measures as an indicator for worsening heart failure status. Researchers identified the Wello health tracking device (Azoi, Inc.) as a target ambulatory health tracker with capacity to track heart rate, blood pressure, and lung capacity. Despite broadly advertised pre-orders and product launch promises, by December 2014 the Wello device failed to ship for consumer purchase. Researchers then turned to a true wearable device that incorporated resting heart rate measurements using a bioimpedance sensor, the Jawbone Up3 (Jawbone, Inc.). After a several months' long delay in delivery, researchers were able to obtain Jawbone Up3 devices with bioimpedance-based resting heart rate measures for the clinical validation study.

3.2.2 Methods

Before recruiting study participants, researchers wore the Jawbone Up3 activity monitors for several weeks to identify challenges with the hardware, web interface, and smartphone app in order to better manage the research participant experience. Researchers identified several issues, including hardware malfunction that caused the device to smoke on charging, challenges with the bracelet clasp design, and battery life shorter than advertised. Researchers used these experiences to design an instruction guide for study participants including instructions for band placement and a prescribed charging schedule.

In addition to hardware and software function, we were concerned about protecting subject privacy in the context of using commercially-available wearable devices that may not employ privacy protections consistent with United States Health Insurance Portability and Accountability Act (HIPAA) laws. We trained patients in the use of the Jawbone Up3 smartphone app and desktop interface, using password protection and data sharing capabilities as available from the manufacturer. Language on the participant consent document included, "The researchers will review your Jawbone Up3 data daily, and will report any abnormal findings to your healthcare provider," and "Efforts will be made to limit the use and disclosure of your personal information, including research study and medical records, to people who have a need to review this information. We cannot promise complete secrecy. Organizations that may inspect and copy your information include the IRB and other representatives of this organization."

Researchers recruited 11 subjects over a 10-month period January 2016 to December 2016. Subjects included patients with an established diagnosis of heart failure who had an implanted cardiac monitor as the gold-standard comparator to validate wearable device findings related to resting heart rate and heart failure status. Subjects uploaded study data to the Jawbone website and shared device data with researchers. Researchers monitored subjects' data to ensure that they were wearing the monitor on a daily basis and particularly at night in order to capture resting heart rate data.

3.2.3 Data analysis

Subject data from the Jawbone were compared to data from the implanted cardiac monitor to assess correlation in resting

heart rate and indicators predicting worsening heart failure. Eight subjects completed the 4-month monitoring period. No subjects reported adverse outcomes. We sought to compare tabular Jawbone data to raw data from implanted devices. However, our clinician partners were unable to obtain raw data from implanted device manufacturers in all but one case. Therefore, we were limited to performing visual correlation of graphic data available from implanted devices to the tabular data from the Jawbone.

In addition to collecting device data, we measured quality of life using validated psychometric scales (Minnesota Living with Heart Failure [MLHF], University of Minnesota; SF-36, RAND Corporation).

3.2.4 Results

Two subjects had extremely poor data correlation with the wearable device failing to generate actionable signals to indicate impending worsening of heart failure status. Six subjects had adequate data capture from the wearable device with reasonable correlation to implantable monitors. However, the majority of the measures clustered around 60 beats per minute, correlating to the programmed lower pacing rate for most study participants.

Six of 8 subjects had clinically significant score improvement on MLHF (>5-point increase). Paired t-test revealed with statistically significant improvement in physical subscore ($p = .03$) and nonsignificant improvement on emotional subscore and overall score. SF-36 demonstrated statistically significant improvement in emotional wellbeing subscore ($p = .03$) with nonsignificant changes in other subscores including increases in physical functional, role physical, and pain subscores, as well as general health score. Despite significant improvements in emotional wellbeing subscores, we noted decline in role emotional, energy, and social subscores (Tables 8 and 9).

3.2.5 Study summary

Conclusions from this pilot study suggested that patients with chronic heart failure can use a wearable monitor without adverse effects, though the utility of the monitor for precise tracking of resting heart rate to predict heart failure status is limited. However, there may be some improvement in quality of life for patients using wearable monitors to measure chronic health status. The validation component of the study failed to demonstrate that the Jawbone Up3 consumer-grade wearable device could be a reliable resting heart rate monitor for patients with heart failure. However, the findings of this study were limited by the fact that most of the study participants had a basement heart rate of 60 beats per minute with chronic pacing, limiting useful resting heart rate data for many participants.

Additional observations related to study outcomes noted challenges related to the lack of funding directed to the clinical site. Without funding incentives for the clinical site, clinical staff were modestly helpful with subject recruitment but otherwise minimally involved with the study, including no engagement in device management and data acquisition. Device-related responsibilities fell to the graduate research assistant (employed at 20 h per week) and highlights the challenge with integrating new technology into clinical practice in terms of workforce sustainability.

Regarding data privacy, we noted that no study participants expressed concerns about privacy protections at the time of recruitment, study consent, or at any point during the study period. We noted at around the same time we were conducting this study in 2014–2015, the Connected & Open Research Ethics (CORE) Project (https://thecore.ucsd.edu) was taking shape to address ethical research with sensor-based devices.

Table 8 Minnesota living with heart failure results

	Mean difference pre- to post-test (range)	p value (* = significant)
Total score	−10.625 (−46 to +14)	.07
Physical subscore	−6.875 (−23 to +4)	.03*
Emotional subscore	−2.25 (−9 to +4)	.12

Table 9 SF-36 results

	Mean difference pre- to post-test (range)	p value
General Health	.625 (−25 to +30)	.46
Physical Functional	8.75 (−20 to +35)	.09
Role Physical	5.21 (−25 to +67)	.32
Role Emotional	−16.67 (−67 to 0)	.10
Energy	−1.88 (−20 to +20)	.35
Emotional Wellbeing	7.17 (−8 to +24)	.03*
Social	−9.06 (−48 to +13)	.14
Pain	9.69 (−30 to +55)	.16

4 Discussion

The Project Honeybee observational clinical trial experience yielded several successes and highlighted some significant challenges. Some issues were specific to our experience, while others may be extrapolatable to other project settings.

4.1 Successes

4.1.1 Nursing students

The ASU doctoral nursing students were an inspiration for us all. They made things work under the most trying of

circumstances. They are highly dedicated, entrepreneurial and problem solvers. They universally reported the experience they had with Project Honeybee was one of their most important learning opportunities of their careers. They explained that while the trials were often compromised through low patient numbers, faulty devices, poor patient compliance, etc. the experience gained in overcoming unanticipated challenges would allow them to do larger and more successful studies with wearable devices in the future.

4.1.2 Publications and presentations

As of this writing, investigators had 9 manuscripts submitted or accepted to peer-reviewed journals, and 10 conference posters or papers delivered.

4.1.3 Faculty and clinician engagement

One of our overarching goals for the Project Honeybee experience was to engage ASU research faculty with area clinicians to establish or build on existing relationships for clinical-based studies. Over the four years of Project Honeybee funded studies, 12 ASU faculty and 39 area clinicians participated in one or more projects.

4.1.4 Clinical institutional engagement

Two of the eight design aspirations of Arizona State University's charter as a New American University are to conduct use-inspired research with purpose and impact and be socially embedded by connecting with communities through mutually beneficial partnerships. In line with these principles, Project Honeybee studies engaged participation with nine external healthcare institutions, including private practice groups, large tertiary medical centers, and a federally qualified healthcare center.

4.2 Challenges

4.2.1 Devices

Although hundreds of commercial ambulatory and wearable devices are on the market, device performance was the major limitation in nearly all trials. First, they are nearly all meant for the health enthusiast consumer market and have not been validated for the intended application or patient population. Device validation is the most important deficiency in the field and any team intending to obtain reliable data must go through a validation process with their own patient population. Second, algorithms for analyzing data are proprietary and hence it is usually impossible to reanalyze the data for key signals. Third, device memory is often a limitation and many devices cannot store the amount of data needed. Fourth, data downloads for many devices require that the device be brought in by the patient. Fifth, many devices advertised are not yet available and plans built on their availability are often frustrated. Sixth, many devices are not comfortable to wear, and patient compliance can be compromised.

4.2.2 IRB approval

ASU's Institutional Review Board does not provide oversight for clinical trials, even if only observational. Therefore, IRB approval was typically obtained through the clinical site of the trial with ASU's IRB providing secondary oversight for ASU-based investigators. Although the ASU IRB has developed relationships with other area IRBs, this process proved quite cumbersome for investigators. Typically, teams underestimated the time required for IRB approval and this often delayed the onset of trials by several months. We began withholding awarded funds until IRB approval was obtained.

4.2.3 Institutional affiliations

Permission for ASU doctoral nursing students to enroll and monitor patients at clinical sites required a formal approval process at the clinical site. Some sites, especially Mayo Clinic, were slow to provide authorization. This factor delayed the onset of some trials.

4.2.4 Patient enrollment

Teams frequently overestimated the availability of patients at a clinical site for the trial. Not all patients are willing to participate, satisfy exclusion criteria and are compliant with study needs. Moreover, on-site recruitment support from clinicians and staff, particularly at non-research institutions, was often lacking. In future trials a comprehensive assessment of patient numbers and on-site subject recruitment support is essential before a trial is launched.

4.2.5 Leadership

Clinicians are very busy and were not reimbursed in our study. Nursing students frequently found it difficult to get guidance from the clinician. In addition, many clinicians were unwilling to devote scarce time to talk with patients about study participation. Coupled with an ASU IRB prohibition for study staff to conduct chart review for potential subject screening, this limitation proved a significant barrier to study subject recruitment. What seemed to work best was to establish a brief weekly meeting between student and clinician. In addition, it would undoubtedly help to reimburse clinicians for their time by incorporating a per-subject enrollment fee as is common in clinical drug and device trials.

5 Conclusion

The Project Honeybee studies have identified many challenges to the deployment of consumer wearable devices in clinical medicine. The field is still immature, and progress will depend significantly on improvements in devices, their validation, wearability, data storage and access to raw data as well as in their ability to monitor several physiological parameters at the same time.

In addition, our experience highlighted the need for adequate time and funding in order to successfully complete a clinical project. Although doctoral nursing students provided outstanding research support interfacing with clinical investigators, clinical support staff, and patient subjects, the 12-month timeframe often posed a challenge for recruitment and longitudinal follow-up even for a small-sized study population of 15–30 subjects. Similarly, the funding limitation that excluded covering tuition costs for graduate research assistants, clinical site incentives, or subject incentives may have contributed to the challenges this project experienced.

Looking ahead, we hope that our early experiences with wearable research may inform continued work to understand the applications, potential, and practice implications of incorporating wearable devices into clinical care.

Acknowledgments The authors and all participants wish to acknowledge the generous support from the Strategic Investment Fund, Arizona State University.

Compliance with ethical standards

Ethical approval All procedures performed in studies involving human participants were in accordance with the ethical standards of the institutional and/or national research committee and with the 1964 Helsinki declaration and its later amendments or comparable ethical standards.

Reference

Deurenberg P, Weststrate JA, Seidell JC. Body mass index as a measure of body fatness: age- and sex-specific prediction formulas. Br. J. Nutr. **65**(2):105–114 (1991)